

## 音楽検索のための楽曲特徴量についての調査 A Survey on Music Features for Music Retrieval

張 夢堯<sup>†</sup>, 成 凱<sup>‡</sup>

Mengyao Zhang, Kai Cheng

### 1 はじめに

大規模の音楽データベースから聴きたい楽曲を効率的に探し出すために、音楽検索が重要である。音楽検索には、テキストによる楽曲検索、鼻歌検索、音楽認識など様々な方式が存在する。最も一般的な検索方式は曲名、歌手名、歌詞をキーワードとして入力することで楽曲を検索する方式である。しかし、多くの場合、ユーザが検索したい音楽は、路上を歩いているときに何気なく聞いた曲である。この場合はテキストによる検索が無力になって、鼻歌検索と音楽認識が有効である。鼻歌検索も音楽認識も、メロディ検索が主である[1]。メロディ検索を行う際に、どの楽曲がどのような特徴を持っているかを利用して行う場合が多いため、楽曲の持つ特徴を定量化する楽曲特徴量が必要である。楽曲特徴量が様々あるが、どれが有効であるかについて、まだ体系的に解明されていない。

本研究では、音楽データのメロディ検索に利用できる楽曲特徴量を調査し、体系的に評価を行う。そのうえ、楽曲特徴量抽出の高速化を図る。

### 2 楽曲特徴量

音楽データには、オーディオデータと MIDI データの 2 種類がある。オーディオデータとは、パソコンのマイク/ライン端子から音声を取り込んでデジタルデータに変換したものである。オーディオデータは WAV、MP3 などのフォーマットがあり、音のそのものを記録したデータである。MIDI データとは演奏情報をやりとりするための規格である。MIDI の情報には、MIDI チャンネルという 1 から 16 の番号が割り当てられている。MIDI チャンネルを使って、1 本の MIDI ケーブルで同時に 16 パート用の情報を送る仕組みになる。

楽曲特徴量とは音楽データの特徴・特性を、定量的に表した数値のことである。音楽データは収録条件、個人差、データ差などの違いが大きくて本質的な部分がわかりにくいし、データ量や計算量が多くて取り扱いにくい。生のデータそのものを利用するのは無理がある[2]。より無駄が少なく、データの本質をあらわした良い楽曲特徴量を抽出する必要がある。以下に楽曲特徴量を列挙する。

#### 2.1 オーディオデータに関する特徴量

オーディオデータに関する特徴量には音域、音高差頻度、音高差・音長差パターンの頻度、音長差平均、基本周波数、MFCC、log-mel spectrum などが使われる。

1. MFCC: 人間の低音に敏感で高音に鈍いという特徴を考慮しながら、声のスペクトルの概形を表現する特徴量である。

2. log-mel spectrum: log-mel spectrum は MFCC の離散コサイン変換無いバージョンである。離散コサイン変換のプロセスがないため、log-mel spectrum は MFCC に比べて計算コストが低いし、計算速度が上がる。

#### 2.2 MIDI データに関する特徴量

MIDI データに関する特徴量にはチャンネルタグ、チャンネル号、チャンネルの平衡度、平均パワー、総発音時間、エネルギー面積、非同時音符率などが使われる。

1. チャンネル号: 各チャンネルには、その演奏情報をどのチャンネルに送るかを記録するためのビットがついている。MIDI の 10 番チャンネルは一般に打楽器やドラムとして使われているため、10 番チャンネルは主旋律チャンネルとして排除することができる。

2. チャンネルの平衡度: 通常、主旋律の発音は左右チャンネル音量のバランス位置にある。各チャンネルには音符が左チャンネルと右チャンネルの音量の大きさ割合を表 PAN 値がある。チャンネルバランス位置からずれた大きさを利用して、そのチャンネルが主旋律チャンネルである可能性を判断できる。

3. 平均パワー: 通常、主旋律チャンネルの平均パワーは他のチャンネルよりはるかに高い。したがって、各チャンネルの平均パワーを利用して、そのチャンネルが主旋律チャンネルである可能性を判断できる。

4. 総発音時間: チャンネルの発音時間の長さを定義する特徴量である。一般的に、主旋律は音楽作品の全体を貫く、主旋律チャンネルの総発音時間は一番長いはずであるため、チャンネルの総発音時間が多いほど、主旋律チャンネルとなる可能性が高い。

5. エネルギー面積: メロディ特徴では音符の音長も音高も重要である。主旋律判断を行う際にはこの 2 つの特徴を総合的に考慮すべきである。したがって、エネルギー面積でどれが主旋律チャンネルを判断できる。

6. 非同時音符率: MIDI ファイルの研究により、主旋律の音符はいつも独立に発音するが、伴奏の音符は同時に発音することが多いため、非同時発音した音符の数量が多いほど、このチャンネルが主旋律チャンネルとなる可能性が高い[5]。

オーディオデータから音高、音長、テンポ、リズムなどの音の性質に基づいて特徴量を抽出してメロディ検索に利用できる。MIDI データには音楽の全音符や演奏の過程、特徴情報が記録されているため、音楽の主旋律抽出が簡単である。したがって、MIDI データはメロディ検索に応用されている場合が多い[4]。

<sup>†</sup>九州産業大学情報科学研究科 Graduate School of Information Science, Kyushu Sangyo University

<sup>‡</sup>九州産業大学情報科学研究科 Graduate School of Information Science, Kyushu Sangyo University

### 3 音楽検索に適する特徴量の考察

これまで述べてきた楽曲特徴量はいずれもメロディ検索が可能であるが、入力された音声データは1番目の音符から検索の目標曲と一致しなければならない。実際の音楽検索では、入力された音声データは楽曲の中間部分から始まることが多く、楽曲全体のごく一部のみである。音楽検索に適する楽曲特徴量は、音楽検索のこの特性を満たさなければならない[3]。このため、いくつかの楽曲特徴量が適用されなくなっている。例えば、音域という楽曲特徴量を用いて検索を行うと、入力された音声データは楽曲全体のほんの一部であるため、その音域は他の音楽を区別する特徴とすることはできない。

### 4 予備実験

実際に音声特徴量の可用性を検証するために、WAVファイルからMFCC(メル周波数ケプストラム係数)というスペクトル特徴量の抽出実験を行った。MFCCは人間の音高知覚が考慮された周波数を表す特徴量である。音声認識、音楽検索などで使われる。今回はMFCCの抽出手順に従って、音声データを入力しMFCCデータが得られた。言語はPythonである。

#### 4.1 使用データ

本研究では、MIREX2020データセットを実験データとして行う。MIREX2020は第16回音楽情報検索評価コミュニケーションで公開されたデータセットである。

#### 4.2 実験結果

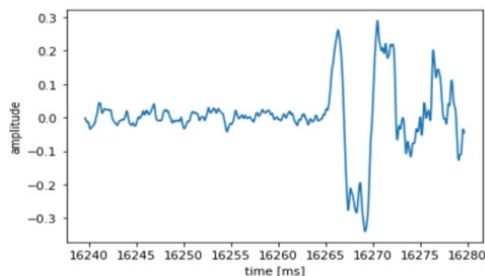


図1: 音響データの音声波形の中心部分

まずは前処理として波形を適当な長さに分割し、窓関数をかけて高速フーリエ変換を行う。

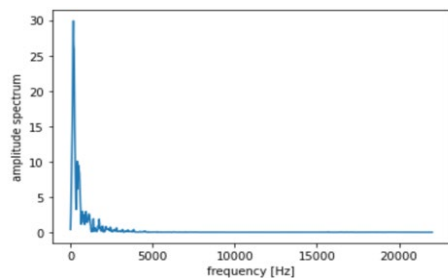


図2: 窓処理を行った振幅スペクトル

元の振幅スペクトルとフィルタバンクをかけた後のスペクトルを比べてみる。

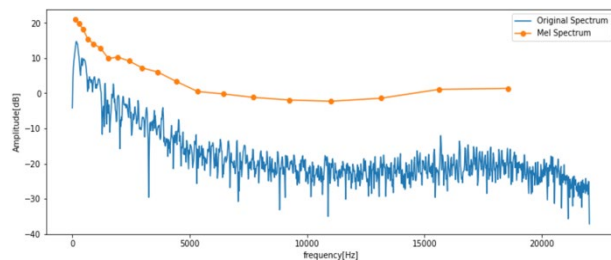


図3: 元の振幅スペクトルとフィルタバンクをかけた後のスペクトル

次は離散コサイン変換を行う。離散コサイン変換の結果から低次の係数を取り出すと、MFCCが得られる。大体12次まで取ることが多い。結果が以下のように出力した。

```
[931.13709989, 1098.67656153, 775.76976108, 520.12292202
383.2883047, 257.18420403, 159.79484762, 91.52241212
60.05333694, 32.27291117, 8.30627297, -9.79333575]
```

### 5 まとめ

音楽検索にはメロディ検索が主な検索手法である。メロディ検索を行う際に利用する楽曲特徴量が数多くある。その中の音楽検索に適する楽曲特徴量を考察し、抽出実験を行った。今後の研究では音楽検索に適する楽曲特徴量どれが有効であるかについて、体系的に評価していく予定である。

#### 参考文献

- [1] 角尾 衣未留, 井上 晃, 西口 正之, 鼻歌検索システムのための楽曲からのボーカルメロディ推定, Vol.2013-MUS-100 No.18
- [2] 奥乃 博, 北原 鉄朗, 吉井 和佳, 楽曲の特徴量抽出と検索技術, 電学誌, 127 巻 7 号, 2007 年
- [3] 市川裕貴, 長名優子, 不応性を有する自己組織化特徴マップを用いた音楽検索, Copyright 2013 Information Processing Society of Japan. All Rights Reserved.
- [4] 胡麻田学, 長名優子, MIDI キーボードによるキーの入力が可能な不応性を有する自己組織化特徴マップを用いた音楽検索システム, Copyright 2012 Information Processing Society of Japan. All Rights Reserved.
- [5] 中山 達喜, 吉田 真一, 音楽分類における特徴量の検討, 26th Fuzzy System Symposium (Hiroshima, September 13-15, 2010)